Werken met FineReader 4.0

Een intelligente kop

Je hebt net een aantal pagina's met tekst doorgefaxt gekregen, maar jammer genoeg staan er fouten in. Het is niet mogelijk om ze nogmaals te laten doorzenden. Ga je dan alle pagina's zelf intikken? Natuurlijk niet! Met het programma FineReader kan je de bladzijden inscannen en naar je tekstverwerker exporteren, om ze daar dan naar hartelust te bewerken...

Er zijn natuurlijk heel wat meer mogelijkheden. Als je om een of andere reden enkele pagina's uit een tijdschrift nodig hebt, kan je deze ook heel gemakkelijk inscannen en door FineReader laten bewerken. Tekst, tabellen en figuren worden allemaal herkend en als dusdanig behandeld. We leggen even kort uit hoe het allemaal in z'n werk gaat...

Stap 1
Installatie

Om FineReader 4.0 van op de CLICKX cd-rom te installeren, moet je je eerst registreren. Start de cd-rom, kies FINEREADER en REGISTREREN en volg de instructies op het scherm. ABBYY stuurt je vervolgens via e-mail een registratienummer toe. Na ingave van dat nummer kan je het programma installeren. Tijdens de installatieprocedure moet je ondermeer je taal ingeven en kiezen welk type van installatie je wil uitvoeren.

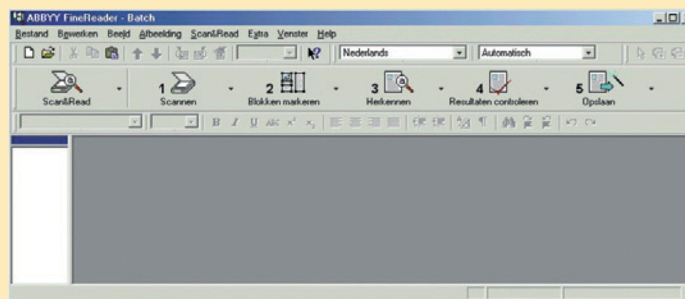


Kies het soort installatie.

Het is aan te raden om te kiezen voor VOLLEDIGE INSTALLATIE. Heb je gebrek aan ruimte op je harde schijf, dan kies je voor AANGEPASTE INSTALLATIE, waar je enkele megabytes aan ruimte kan besparen door herkenning van bijvoorbeeld het Arabisch uit te schakelen. Zorg er bij de aangepaste installatie wél voor dat je de PROGRAMMA-SHELL installeert.

Stap 2
Een eerste scan

Na de installatie is er bij START, PROGRAMMA'S een menu voor ABBYY FineReader toegevoegd. Daarin hebben we de keuze tussen FineReader 4.0 Professional en FineReader 4.0 Forms. We kiezen voor PROFESSIONAL FINEREADER. Het programma wordt opgestart en we krijgen een leeg scherm te zien. Dat spreekt natuurlijk vanzelf, want we hebben nog niets ingescand. Plaats dus de pagina die je wil inscannen op de scanner en klik op SCAN & READ linksboven op het scherm.



Het openingsscherm van FineReader.

Er wordt een wizard opgestart die je doorheen het proces zal leiden. Wordt de wizard tijdens één van de stappen om een of andere reden onderbroken, dan kan je de vijf afzonderlijke stappen ook zelf uitvoeren met behulp van de knoppen die je op de taakbalk bovenaan vindt.

Indien er meerdere scanners op het systeem geïnstalleerd zijn, vraagt FineReader welke scanner gebruikt moet worden. Selecteer de gewenste scanner en klik op VOLGENDE. FineReader zal nu je scanprogramma openen. Hier gaan we niet dieper op in, omdat dit scherm voor elke scanner verschillend is. Let wel op dat je inscant met een *resolutie* die tussen 300 en 600 *dpi* ligt, om de tekstherkenning goed te laten verlopen. Na het inscannen vraagt de wizard je naar de taal waarin het document is opgesteld.



Een Nederlandstalig document.

Kies de juiste taal en klik weer op VOLGENDE. Nu gaat FineReader pas echt aan het werk. De gescande pagina wordt geanalyseerd. Tekst, kolommen en afbeeldingen worden herkend en gedefinieerd.

ieermachine





ALS JE OCR ABSOLUUT NIET WERKT...

Links bovenaan zien we de gescande pagina geïnterpreteerd als een beeld. Dit scherm verschilt dus niet zoveel van het voorbeeldscherm dat te zien is als we een pagina aan het inscannen zijn. De verschillende bloktypes hebben een volgnummer gekregen en zijn van elkaar te onderscheiden door een ander kadertje. Tekst wordt omkaderd in het groen, figuren hebben een rood randje gekregen, tabellen worden in het bordeaux omlijnd.

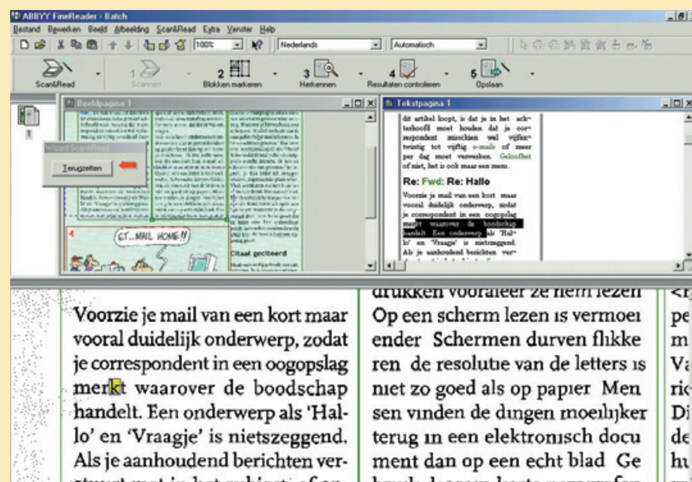
Indien we dat wensen, kunnen we een blok van type doen veranderen. Selecteer het gewenste blok, klik op de rechtermuisknop en kies bij **Bloktype** voor **tekst**, **tabel**, **illustratie** of **barcode**. Je kan FineReader het blok ook laten negeren.

Onderaan op het scherm kunnen we inzoomen op de beeldpagina. Klik hiervoor op de rechtermuisknop en kies voor **Schaal**. Rechts bovenaan is het tekstvenster. Daarin wordt het resultaat van de geopende pagina weergegeven. Je zal zien dat de lay-out er niet zo uitziet als op het brondocument. Alle blokken worden gewoon onder elkaar geplaatst. Dit scherm is niet *WYSIWYG*, maar de lay-out zal uiteindelijk wel hetzelfde zijn als op het origineel. Dat zal je wel merken als we het document exporteren naar Word.

Indien je de pagina per ongeluk ondersteboven hebt ingescand, wordt dit prompt gecorrigeerd. Alle tekst wordt dan achtereenvolgens vertaald naar regels, tekens, cijfers en/of letters. De wizard meldt dan met trots dat de herkenning voltooid is. Dit is niet zonder voorbehoud, we worden verondersteld het document nog te controleren op fouten.

Stap 3
Controleren op fouten

We klikken op **Minimaliseren** en zien dat het scherm in drie delen is opgesplitst.



Het hoofdscherm van FineReader.

Als we in het tekstvenster iets selecteren, wordt dat ook zichtbaar in het zoom-venster onderaan. De bedoeling van dit tekstvenster is om eventuele herkenningsfouten van FineReader te corrigeren. Alles waar FineReader zijn twijfels over heeft, is in een groen kleurtje gestoken.

Overloop de tekst en pas aan waar nodig. Je kan hier eventueel ook tekstblokken weglaten.

Als we verder willen gaan, klikken we op **Terugzetten**, waarmee we onze wizard opnieuw tevoorschijn toveren. Als we nu kiezen voor **Goed, maar sommige fouten dienen gecorrigeerd te worden**, dan treedt de spellingcontrole in werking. Als je zeker bent dat je alle fouten er zelf hebt uitgehaald, kies je voor de optie **Een succes, geen fouten bespeurd.**



Resultaat controleren.

Wij kiezen echter voor de zekerheid en opteren voor de extra controle. Klik op **Volgende** en vervolgens op de knop **Resultaten controleren** die je in de taakbalk bovenaan kan vinden. FineReader doorloopt nogmaals het document en zal je - net zoals een gewo-

ne spellingcontrole - bij een woord dat het niet kent, vragen wat het moet doen. Je kan het woord negeren, vervangen of aanpassen.

Stap 4
De scan opslaan

Na de spellingcontrole wil FineReader weten of je wil verdergaan met pagina's inscannen of dat je het document wil exporteren/opslaan. We kiezen voor **Exporteren/Opslaan** en klikken op **Volgende**.



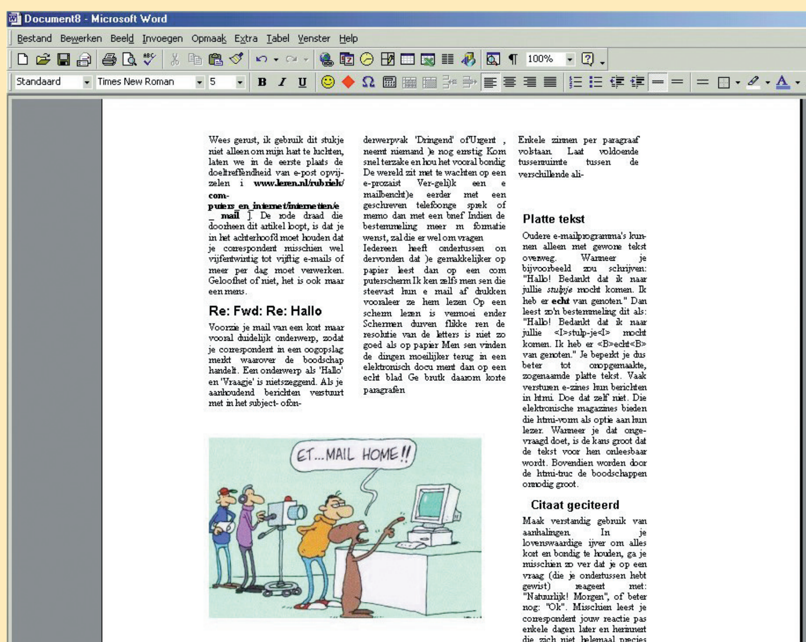
Scannen of opslaan?



Wat willen we met de scan doen?

Nu kunnen we kiezen naar welk programma we de scan willen exporteren. We hebben de keuze tussen Excel, Word, het klembord, e-mail of de scan opslaan naar een bestand.
We kiezen voor Word, vinken het bolletje bij **Alleen geselecteerde pagina's opslaan** aan, en klikken op **Voltooien**. Worden alle blokken nog steeds onder elkaar weergegeven, kies dan bij **Beeld** voor **Afdrukweergave**. Speel nog wat met de lengte en breedte van de kolommen, verplaats ze een beetje indien nodig en je zou hier zonder moeite een vrijwel identieke lay-out als de originele pagina moeten krijgen.

Kiezen we bij het opslaan voor **e-mail**, dan wordt je standaard e-mailprogramma geopend met daarin de scan als bijlage. Deze bijlage wordt opgeslagen als RTF-bestand (Rich Text Format).
Bij **Opslaan naar bestand** kunnen we ondermeer kiezen voor PDF-, HTML- en DBF-bestanden.



Document8 - Microsoft Word

Wees gerust, ik gebruik dit stukje niet alleen om mijn hart te luchten, laten we in de eerste plaats de doeltreffendheid van e-post opvijzelen i **www.leren.nl/rubriek/computers_en_internet/internetten/e_mail**]. De rode draad die doorheen dit artikel loopt, is dat je in het achterhoofd moet houden dat je correspondent misschien wel vijfentwintig tot vijftig e-mails of meer per dag moet verwerken. Geloofhet of niet, het is ook maar een mens.

Re: Fwd: Re: Hallo

Voorzie je mail van een kort maar vooral duidelijk onderwerp, zodat je correspondent in een oogopslag merkt waarover de boodschap handelt. Een onderwerp als 'Hallo' en 'Vraagje' is nietszeggend. Als je aanhoudend berichten verstuurt met in het subject- ofonderwerpvak 'Dringend' of'Urgent , neemt niemand)e nog ernstig Kom snel terzake en hou het vooral bondig De wereld zit met te wachten op een e-prozaist Ver-gelijk een e mailbericht)e eerder met een geschreven telefoonge sprek of memo dan met een brief Indien de bestemmeling meer in formatie wenst, zal die er wel om vragen
Iedereen heeft ondertussen on dervonden dat)e gemakkelijker op papier leest dan op een com puterscherm Ik ken zelfs men sen die steevast hun e mail af drukken vooraleer ze hem lezen Op een scherm lezen is vermoei ender Schermen durven flikke ren de resolutie van de letters is niet zo goed als op papier Men sen vinden de dingen moeilijker terug in een elektronisch docu ment dan op een echt blad Ge bruik daarom korte paragrafen

Enkele zinnen per paragraaf volstaan. Laat voldoende tussenruimte tussen de verschillende ali-

Platte tekst

Oudere e-mailprogramma's kunnen alleen met gewone tekst overweg. Wanneer je bijvoorbeeld zou schrijven: "Hallo! Bedankt dat ik naar jullie *stukje* mocht komen. Ik heb er **echt** van genoten." Dan leest zo'n bestemmeling dit als: "Hallo! Bedankt dat ik naar jullie <I>stulp-je<I> mocht komen. Ik heb er <B>echt<B> van genoten." Je beperkt je dus beter tot onopgemaakte, zogenaamde platte tekst. Vaak versturen e-zines hun berichten in html. Doe dat zelf niet. Die elektronische magazines bieden die html-vorm als optie aan hun lezer. Wanneer je dat ongevraagd doet, is de kans groot dat de tekst voor hen onleesbaar wordt. Bovendien worden door de html-truc de boodschappen onnodig groot.

Citaat geciteerd

Maak verstandig gebruik van aanhalingen. In je lovenswaardige ijver om alles kort en bondig te houden, ga je misschien zo ver dat je op een vraag (die je ondertussen hebt gewist) reageert met: "Natuurlijk! Morgen", of beter nog: "Ok". Misschien leest je correspondent jouw reactie pas enkele dagen later en herinnert die zich niet helemaal precies

Het resultaat in Word.

Wil je na het exporteren de originele scan ook bewaren, dan kan dit door een **batch** aan te maken. Een batch is de gegevensopslagplaats van FineReader. De gescande afbeeldingen, de herkende tekst en andere gegevens zoals gebruikerspatronen en talen worden in de batch opgeslagen.
Je maakt een batch aan door **Bestand** en dan **Batch opslaan** te kiezen. Geef een naam in voor je batch en zeg waar de bestanden bewaard moeten worden.
Klik vervolgens op **Voltooien** en FineReader zal al de informatie (waaronder een jpg-bestand voor elke gescande pagina) in de geselecteerde map opslaan.



Opslaan van je scans.

Stap 5
Bewerk een bestaande figuur

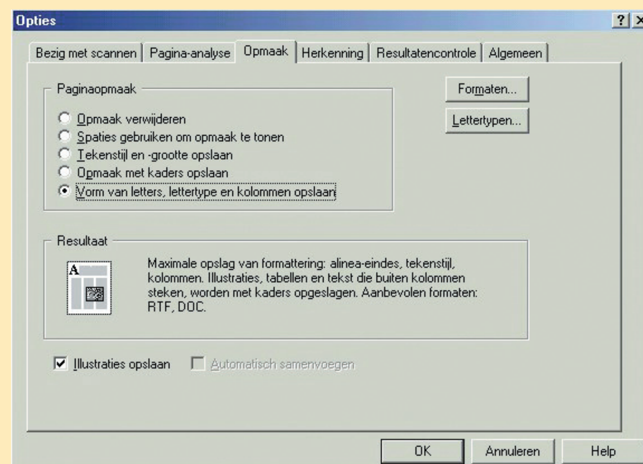
Als je een afbeelding wil bewerken die je al op je harde schijf hebt staan, dan kan dat natuurlijk ook. Klik op het pijltje rechts van **Scan&Read** en een klein keuzemenu komt tevoorschijn. Selecteer **Open&Read** en je kan het bestand dat je wil bewerken aanduiden.



Een afbeelding bewerken.

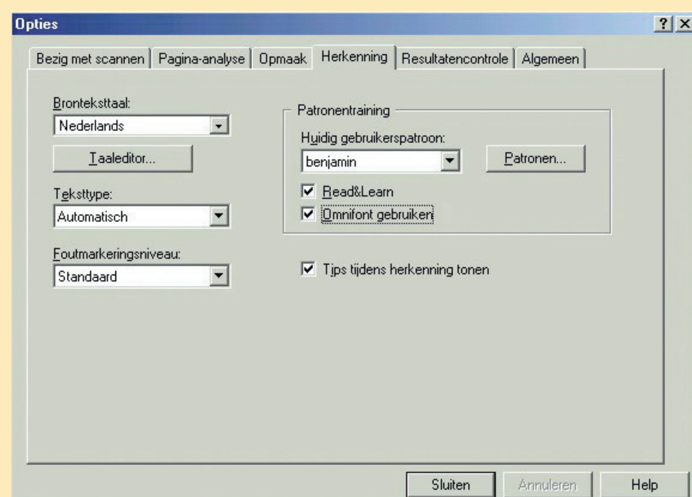
Stap 6
Opties Instellen

In FineReader kan je een heleboel opties instellen om een zo goed mogelijk resultaat te bereiken. Soms kan het zijn dat je al die ta-



Alleen tekst nodig?

bellen en opmaakstijlen niet nodig hebt. Dat je enkel tekst wil, zonder de afbeeldingen erbij. Dan kies je bij **Extra** voor **Opties**, en daarin voor het tabblad **Opmaak**. Bij **Paginaopmaak** kan je kiezen welke mate van formattering er van je document opgeslagen moet worden. Onderaan het scherm kan je kiezen om illustraties niet op te slaan.



Stel een gebruikerspatroon in.

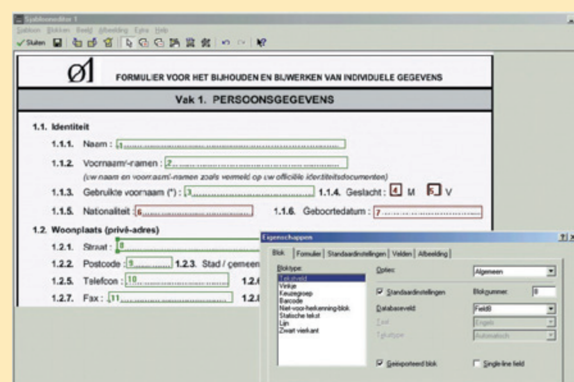
Onder het tabblad **Pagina-analyse** kan je instellingen voor tabellen en kolommen opgeven, bij **Herkenning** kan je ondermeer de strengheid van de foutmarkering opgeven. Een **Gebruikerspatroon** is vooral handig indien het gaat om teksten in decoratieve lettertypes of teksten die een hoop speciale symbolen (zoals wiskundige formules) bevatten.
Bij het trainen van een patroon moet je eerst enkele pagina's laten herkennen in de trainingsmodus. Je zal zien dat FineReader je bij een heleboel tekens de correcte taal en het juiste symbool zal vragen.
Vervolgens zal FineReader dit patroon gebruiken als hulpmiddel bij het inscannen van de rest van de tekst. Het spreekt vanzelf dat de extra tijd die je hierin moet steken niet rendabel is bij het scannen van pakweg vier pagina's.



Train FineReader met een gebruikerspatroon.

Stap 7
Sjablonen

Als je een groot aantal documenten met dezelfde indeling - zoals formulieren of vragenlijsten - gaat verwerken, dan kan het handig zijn een sjabloon aan te maken. Een sjabloon bestaat uit een aantal blokken van een bepaald type die je gebruikt om de pagina in te delen.
Dat sjabloon kan je dan toepassen op alle pagina's die deze indeling hebben. Wil je een sjabloon aanmaken, klik dan vervolgens op **Start**, **Programma's**, **ABBYY FineReader** en **FineReader Forms**. We kiezen voor een **Nieuwe batch aanmaken**, vervolgens voor een **nieuw sjabloon** dat we zullen verkrijgen door een afbeelding te **scannen**. Na het scannen krijgen we de **sjablooneditor** te zien. Daarin kan je op je scan blokken aanduiden en definiëren.
Elk blok kan je een naam geven, je kan ook zeggen of het gaat om een tekstvak of bijvoorbeeld een hokje om aan te kruisen. Velden kan je ook linken aan een database.
Heb je een sjabloon aangemaakt, dan kan je het op een scan toepassen door op de knop **Sjabloon laten passen** te klikken. De andere stappen zijn ongeveer hetzelfde als in de FineReader Professional, alleen kan je hier je resultaten niet laten controleren.



Maak een sjabloon.

Wil je nog meer informatie over het maken van formulieren, bekijk dan eens het voorbeeld dat FineReader bij het programma geleverd heeft. Kies voor **Bestaande batch openen** en selecteer vervolgens het bestand **demo.frm** in de map **\ABBYY FineReader Professional\Demo.**

De site van FineReader vind je op [**www.abbyy.com**], meer informatie over OCR - de techniek die tekens herkent als tekst - op [**www.webopedia.com/TERM/o/optical_character_recognition.html**].

— Benjamin Carlier —

VAKTAAL

Dpi: De afkorting van dots per inch. Letterlijk vertaald: aantal punten per duim (een duim is een Engelse lengtemaat en komt overeen met 2,54 cm). Hiermee wordt de resolutie van een printer (of een scanner) uitgedrukt. Een toestel met een resolutie van 600 bij 600 dpi drukt dus zowel horizontaal als verticaal 600 punten over een afstand van 2,54 cm. Hoe hoger de resolutie, hoe beter de kwaliteit van de afdruk.

Resolutie: Een term om de kwaliteit en scherpte van een beeld, een foto of een afdruk uit te drukken.

RTF: Rich Text Format. Bestand wordt opgeslagen als tekst, maar met behoud van de opmaak.

WYSIWYG: What You See Is What You Get. Een term die gebruikt wordt als het document dat je op je scherm ziet qua lay-out precies hetzelfde is als de afdruk ervan.